AWARD NUMBER: W81XWH-13-1-0348

TITLE: "Screening for Novel Germline Rare Mutations Associated with Aggressive Prostate Cancer"

PRINCIPAL INVESTIGATOR: Zhuo Chen

CONTRACTING ORGANIZATION: Wake Forest University
Winston Salem, NC 27157

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| October 2014 | Annual | 15 Sep 2013 - 14 Sep 2014 |

**4. TITLE AND SUBTITLE**
"Screening for Novel Germline Rare Mutations Associated with Aggressive Prostate Cancer"

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER W81XWH-13-1-0348**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Zhuo Chen

E-Mail: zhchen@wakehealth.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Wake Forest University

Winston-Salem, NC 27157

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Prostate cancer is the most common noncutaneous cancer in males in the U.S. While the major indolent form may not even require treatment, about 10-15% of PCa cases of aggressive form requiring intensive treatment. However, our inability to reliably distinguish between aggressive and indolent PCa early on in the course of the disease has resulted in the over-treatment of many and under treatment of some. We hypothesize that inherited sequence variants in the genome are associated with PCa aggressiveness. Currently, we conducted single variant analysis and gene-based analysis to identify rare variants that have strong effects on aggressive PCa risk in exome-array data among a total of 1,902 PCa cases of European descent, including 464 aggressive PCa cases and 1,438 indolent PCa cases. We successfully identified 11 novel variants associated with PCa aggressiveness in Caucasians and further confirmation in additional Caucasians and African American men are to be conducted. Our study will provide more insight into the etiology of aggressive PCa and provide potential effective targets for therapy of aggressive PCa.

**15. SUBJECT TERMS**
prostate cancer, aggressiveness, association, rare variants, INPP5D, HINFP, exome

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | USAMRMC |
| Unclassified | Unclassified | Unclassified | UU | 22 | 19b. TELEPHONE NUMBER *(include area code)* |

**Table of Contents**

**Page**

## INTRODUCTION

While most prostate cancer (PCa) patients have an indolent form of the disease that may not even require treatment, about 10-15% of PCa patients have an aggressive form that may progress to metastases and death, thus requiring intensive treatment. Several clinical variables such as PSA levels, Gleason grade, and TNM stage are good predictors for disease with poor clinical outcomes; however, their predictive performance needs to be improved. Our inability to reliably distinguish between these two forms of PCa, early on in the course of the disease has resulted in the over-treatment of many and under treatment of some. Another dilemma is a large difference in PCa risk, especially aggressive PCa, between races. African Americans (AAs) have the world's highest incidence of PCa and are twice as likely, as compared with Caucasians to die of the disease. Inherited markers of aggressive PCa could be used for screening and diagnosis of aggressive PCa at an early stage while reducing over-diagnosis and treatment for others. The overall hypothesis is that inherited sequence variants in the genome are associated with a lethal (aggressive) form of PCa but not indolent PCa, and the difference in these variants between races may contribute to higher incidence of and mortality from aggressive PCa in AA.

In this DOD proposal, we proposed: **1)** To identify novel inherited rare variants (MAF<5%) in the exome that are associated with aggressive but not indolent PCa in a case-control population of 1,000 aggressive and 1,000 indolent PCa patients of European descent from John Hopkins Hospital (JHH); **2)** To replicate the rare variants identified in Aim1 in additional 1,000 aggressive and 1,000 indolent PCa patients of European descent from JHH; and **3)** To evaluate the effect of rare mutations confirmed in Aim2 in an African American (AA) population with 500 aggressive and 500 indolent PCa patients from JHH.

## KEYWORDS

prostate cancer, aggressiveness, association, rare variants, INPP5D, HINFP, exome

## OVERALL PROJECT SUMMARY

***Approved Statement of Work:***

**Aim 1. To identify novel inherited rare variants (MAF<5%) in the exome that are associated with aggressive but not indolent PCa in a case-control population of 1,000 aggressive and 1,000 indolent PCa patients of European descent from John Hopkins Hospital (JHH).**

**Tasks**

**Month 1-4:** Preparation of the study, including regulatory review, IRB approval and other logistical issues.

**Month 4-8:** 1) Genotype 1,000 aggressive and 1,000 indolent PCa patients of European descent from JHH using Illumina Human Exome BeadChip platform. 2) Perform bioinformatics analysis for all nonsynonymous variants on Illumina Human Exome BeadChip by PolyPhen2.

**Month 8-12:** Perform single variant analysis using logistic regression and gene-based analysis by SKAT. Select rare variants with a p-value of 1E-5 in single variant analyses and variants in genes with reach a p-value of 1E-4 in gene-based analyses by SKAT.

**Outcome and deliverables**

We expect to identify 2-10 rare variants with a P-value of 1E-5 based on single variant analysis, and also 2-10 genes with a P-value of 1E-4 based on gene-based analysis by SKAT.

***Detailed report***

Study design modification. In our initial proposal, we proposed to genotype 1,000 aggressive and 1,000 indolent PCa patients of European descent from JHH using Illumina Human Exome BeadChip platform. During month 4-8, we genotyped 791 subjects from JHH study, including142 subjects with aggressive PCa and 635 subjects with non-aggressive PCa using Illumina Human Exome BeadChip platform. In addition, we were able to obtain access to the Exome BeadChip array data for two additional Caucasian populations (Michigan and CAPS) with a total of 328 aggressive PCa cases and 814 indolent PCa cases. Therefore, we also conducted a genome-wide association analysis for rare variants with PCa aggressiveness in those two populations. Compared with our original study design, the new design greatly improved our statistical power since we were able to evaluate the replication results for all the rare variants. We were also able to decrease the number of false positive results by including two more populations to compare the association results.

Study Subjects. Subjects included in the John Hopkins (JHH) study were recruited during Jan. 1999 to Dec. 2008. All of them underwent radical prostatectomy for treatment of prostate cancer. Details of this study have been described in previous publications. In this study, aggressive prostate cancer was defined as: 1) Gleason Score $\geqslant$8; or 2) Gleason Score =7, with the most prevalent pattern being 4; or 3) stage T3b or higher; or 4) involvement of regional lymph nodes; or 5) presence of distant metastasis. Otherwise, the cancers were classified as non-aggressive prostate cancer. In this study, we genotyped 791 subjects from the JHH study, including142 subjects with aggressive PCa and 635 subjects with non-aggressive PCa using Illumina Human Exome BeadChip platform.

The second population included subjects recruited in Sweden from the CAPS study, which were diagnosed from Jul. 2001 and Oct. 2003. Details of this study have been described in previous publications. In the CAPS study, aggressive prostate cancers were defined as: 1) Gleason Score

⩾8; or 2) stage T3 or higher; or 3) involvement of regional lymph nodes; or 4) presence of distant metastasis; or 5) serum PSA >50 ng/ml. Otherwise, the cancers were classified as non-aggressive prostate cancer. In this study, 446 subjects from the CAPS study were genotyped by ExomeArray. Among them, 149 subjects were aggressive prostate cancer patients while 297 patients had a non-aggressive form of the disease.

The third study population included subjects recruited by the University of Michigan. The definition of prostate cancer aggressiveness in the Michigan population was exactly the same as in the JHH study. In this study, 696 subjects from the Michigan study were genotyped using the Human Exome BeadChip platform. Among them, 179 subjects were aggressive prostate cancer patients while 517 subjects had a non-aggressive form of the disease.

Genotyping and Quality Control. Genotyping of samples in the first stage was conducted using the Illumina Human Exome BeadChip at the Center for Cancer Genomics, Wake Forest University School of Medicine. A total of 247,870 genetic variants were included in the ExomeArray. Those polymorphic SNPs were used for sex and an IBS check was performed for all subjects using PLINK software (Purcell 2007). In addition, polymorphic SNPs were also used to estimate the missing rate per individual. In each stage, subjects with a genotyping missing rate >5% were removed from further analysis. For subjects in stage 1 with exome data available, an IBS check and sex check were also performed. SNPs with a missing rate >2% in subjects passed quality control (QC) were removed from further analysis.

*Bioinformatics analysis (Variant effect prediction).* All coding nonsynonymous variants were assessed for potential effect by Polymorphism Phenotyping version 2 (PolyPhen2), which is a tool for predicting the possible impact of an amino acid substitution on the structure and function of a human protein. For a given variant, PolyPhen2 calculates a Naïve Bayes posterior probability that the mutation is damaging and then appraised qualitatively as benign, possibly damaging, or probably damaging (Adzhubei 2010).

Statistical analysis for single SNP effect. Principal components analysis was conducted to detect potential population stratification by EIGENSTRAT software (Price 2006). The top 5 eigenvectors which indicates ancestral heterogeneity within a group of individuals were adjusted as covariates in multivariate logistic regression analysis.

All polymorphic genetic variants that passed QC were evaluated for associations with prostate cancer aggressiveness. For genetic variants with any of the genotype counts ⩽5, Fisher’ s exact test was applied to investigate potential association. For genetic variants with genotype counts >5, multivariate logistic regression analysis was conducted assuming an additive genetic model, adjusting for age-at-diagnosis and the top 5 eigenvectors. All analyses were performed using the PLINK software package (Purcell 2007).

Gene-based analysis. We used a novel statistical approach called Sequence Kernel Association Test (SKAT), to conduct gene-based analysis of rare variants for aggressive PCa. SKAT is a supervised and flexible regression method to test for association between rare variants in a gene or genetic region and a continuous or dichotomous trait. Compared to other methods of estimating the joint effect of a subset of SNPs, SKAT is able to deal with variants that have different direction and magnitude of effects, and allows for covariate adjustment (Wu 2011). In addition, SKAT can also avoid arbitrary selection of threshold in burden test. Moreover, SKAT is computationally efficient, compared to a permutation test, making it feasible to analyze the large dataset in our study.

## Results

Detailed clinical and demographic characteristics for the study populations were presented in Table 1.

Table 1. Clinical and Demographic Characteristics of Subjects in Stage 1.

| | JHH # (%) | | MI # (%) | | CAPS # (%) | |
|---|---|---|---|---|---|---|
| Characteristics | Agg (N=142) | Non-Agg (N=635) | Agg (N=179) | Non-Agg (N=517) | Agg (N=149) | Non-Agg (N=297) |
| *Age at enrollment (Year)* | | | | | | |
| Mean (sd) | 51.5 (3.9) | 49.29 (4.44) | NA | NA | NA | NA |
| Age at diagnosis | | | | | | |
| ≤ 55 | NA | NA | 178 (99.4) | 517 (100) | 48 (32.2) | 93 (31.3) |
| > 55 | NA | NA | 1 (0.6) | 0 | 101 (67.8) | 204 (68.7) |
| Missing | NA | NA | 0 | 0 | 0 | 0 |
| *Family History (first-degree relatives)* | | | | | | |
| No | 125 (88.0) | 551(86.8) | NA | NA | 105 (70.5) | 184 (62.0) |
| Yes | 15 (10.6) | 66 (10.4) | NA | NA | 41 (27.5) | 109 (36.7) |
| Missing | 2 (1.4) | 18 (2.8) | NA | NA | 3 (2.0) | 4 (1.3) |
| *PSA levels at diagnosis for cases or at enrollment for controls (ng/ml)* | | | | | | |
| ≤ 4 | 21 (14.8) | 224 (35.3) | 12 (6.7) | 164 (31.7) | 7 (4.7) | 60 (20.2) |
| 4.01-9.99 | 78 (54.9) | 357 (56.2) | 89 (49.7) | 281 (54.4) | 25 (16.8) | 157 (52.9) |
| 10-19.99 | 23 (16.2) | 45 (7.1) | 30 (16.8) | 39 (7.5) | 22 (14.8) | 55 (18.5) |
| 20-49.99 | 18 (12.7) | 4 (0.6) | 20 (11.2) | 4 (0.8) | 25 (16.8) | 23 (7.7) |
| 50-99.99 | 0 | 0 | 19 (10.6) | 1 (0.2) | 25 (16.8) | 0 |
| ≥100 | 0 | 0 | 0 | 0 | 43 (28.9) | 0 |
| Missing | 2 (1.4) | 5 (0.8) | 9 (5.0) | 28 (5.4) | 2 (1.3) | 2 (0.7) |
| *T-stage* | | | | | | |
| T1 | 0 | 0 | 0 | 1 (0.2) | 20 (13.4) | 173 (58.2) |
| T2 | 47 (33.1) | 512 (80.6) | 71 (39.7) | 467 (90.3) | 26 (17.4) | 122 (41.1) |
| T3a | 53 (37.3) | 123 (19.4) | 33 (18.4) | 49 (9.5) | 0 | 0 |
| T3b | 41 (28.9) | 0 | 33 (18.4) | 0 | 0 | 0 |
| T3c | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| T3x | 1 (0.7) | 0 | 0 | 0 | 83 (55.7) | 0 |
| T4 | 0 | 0 | 3 | 0 | 18 (12.1) | 0 |
| TX | 0 | 0 | 0 | 0 | 0 | 0 |
| Missing | 0 | 0 | 39 (21.8) | 0 | 2 (1.3) | 2 (0.7) |
| *N-stage* | | | | | | |
| N0 | 119 (83.8) | 627 (98.7) | 119 (66.5) | 410 (79.3) | 36 (24.2) | 60 (20.2) |
| N1 | 16 (11.3) | 0 | 26 (14.5) | 0 | 22 (14.8) | 0 |
| NX | 1 (0.7) | 8 (1.3) | 20 (11.2) | 107 (20.7) | 91 (61.1) | 237 (79.8) |
| Missing | 0 | 0 | 14 (7.8) | 0 | 0 | 0 |
| *M-stage* | | | | | | |
| M0 | 0 | 0 | 81 (45.3) | 257 (49.7) | 76 (51.0) | 110 (37.0) |
| M1 | 0 | 0 | 15 (8.4) | 0 | 45 (30.2) | 0 |
| MX | 142 (100) | 635 (100) | 72 (40.2) | 260 (50.3) | 28 (18.8) | 187 (63.0) |
| Missing | 0 | 0 | 11 (6.1) | 0 | 0 | 0 |
| *Gleason (biopsy)* | | | | | | |
| ≤ 4 | 0 | 0 | 0 | 6 (1.2) | 0 | 21 (6.7) |
| 5 | 0 | 8 (1.3) | 0 | 21 (4.1) | 9 (6.0) | 49 (16.5) |
| 6 | 1 (0.7) | 420 (66.1) | 6 (3.4) | 272 (52.6) | 25 (16.8) | 163 (54.9) |
| 7 (3+4) | 16 (11.3) | 207 (32.6) | 16 (8.9) | 218 (42.2) | 0 | 60 (20.2) |
| 7 (4+3) | 75 (52.8) | 0 | 84 (46.9) | 0 | 48 (32.2) | 0 |
| 7 (total) | 91 (64.1) | 207 (32.9) | 100 (55.9) | 218 (42.2) | 48 (32.2) | 60 (20.2) |
| 8 | 31 (21.8) | 0 | 31 (17.3) | 0 | 22 (14.8) | 0 |
| 9 | 19 (13.4) | 0 | 35 (19.6) | 0 | 31 (20.8) | 0 |
| 10 | 0 | 0 | 3 (1.7) | 0 | 3 (2.0) | 0 |
| Missing | 0 | 0 | 0 | 0 | 11 (7.4) | 4 (1.3) |

A total of 247,870 genetic variants were included in this ExomeArray. Among them, 92,173, 88,087 and 71,435 genetic variants were polymorphic in JHH, Michigan and CAPS population, respectively. For polymorphic genetic variants, only those with a missing rate >0.98 in subjects passed QC were kept for further statistical analyses, including 91,998 variants in JHH, 87,879 variants in MI and 71,220 variants in CAPS. 79,729, 60,243, 57,126 genetic variants had an MAF < 0.1 in the JHH, Michigan and CAPS population, respectively.

Association Analysis for single variant

We did not observe any association between genetic variants and PCa aggressiveness achieved genome-wide significance (p-value < 5E-3) in JHH, Michigan or CAPS populations. In the JHH population, 47 variants were significantly associated with PCa aggressiveness with a p-value < 1E-3, including 13 rare variants with minor allele frequency (MAF) < 0.05, and 34 common ones (MAF $\geqslant$ 0.05) (Table 2). In the Michigan population, we found 27 variants significantly associated with PCa aggressiveness (p-value < 1E-3), including 11 rare ones and 16 common ones (Table 3). In the CAPS population, we identified 18 variants significantly associated with

PCa aggressiveness (p-value < 1E-3), including 7 rare ones and 11 common ones (Table 4). No variants were significantly associated with PCa aggressiveness with p-value < 1E-3 in all three populations.

Table 2. Associations between genetic variants and PCa aggressiveness in the JHH population with p-value < 1E-3.

| SNP | CHR | BP | Gene | A1/A2 | MAF Agg | MAF Non-Agg | Genotype Agg | Genotype Non-Agg | OR | P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| rs16830693 | 1 | 43,805,240 | MPL | G/A | 0.060 | 0.021 | 1/15/120 | 0/26/599 | 3.14 | 7.62E-04 |
| rs61818256 | 1 | 201,294,910 | PKP1 | A/G | 0.059 | 0.018 | 0/16/120 | 0/22/602 | 3.48 | 3.67E-04 |
| rs17851681 | 1 | 227,954,677 | SNAP47 | A/G | 0.048 | 0.113 | 0/13/123 | 6/129/490 | 0.39 | 7.97E-04 |
| rs17228441 | 2 | 186,627,943 | FSIP2 | A/G | 0.603 | 0.447 | 48/68/20 | 119/319/185 | 1.88 | 1.10E-05 |
| rs992822 | 2 | 186,654,592 | FSIP2 | A/G | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs17229201 | 2 | 186,656,956 | FSIP2 | A/G | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs60029104 | 2 | 186,658,056 | FSIP2 | G/A | 0.613 | 0.455 | 50/66/20 | 126/317/182 | 1.85 | 1.53E-05 |
| rs10490391 | 2 | 186,658,438 | FSIP2 | G/A | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs10490392 | 2 | 186,658,565 | FSIP2 | C/A | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs2161036 | 2 | 186,659,359 | FSIP2 | C/A | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs10931200 | 2 | 186,664,963 | FSIP2 | C/A | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs36004074 | 2 | 186,665,432 | FSIP2 | G/A | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs11695215 | 2 | 186,665,824 | FSIP2 | A/G | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs7605884 | 2 | 186,667,121 | FSIP2 | A/G | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs16827154 | 2 | 186,670,780 | FSIP2 | T/A | 0.603 | 0.453 | 48/68/20 | 120/325/179 | 1.87 | 1.66E-05 |
| rs17826534 | 2 | 186,671,357 | FSIP2 | G/A | 0.603 | 0.448 | 48/68/20 | 119/322/184 | 1.88 | 1.15E-05 |
| rs1862066 | 2 | 186,671,912 | FSIP2 | G/A | 0.599 | 0.447 | 47/69/20 | 118/323/184 | 1.87 | 1.57E-05 |
| rs10804178 | 2 | 210,849,283 | UNC80 | G/A | 0.401 | 0.512 | 23/63/50 | 153/334/138 | 0.59 | 2.52E-04 |
| rs61729839 | 2 | 238,277,379 | COL6A3 | A/G | 0.044 | 0.011 | 0/12/124 | 0/14/611 | 4.08 | 7.50E-04 |
| bs2_242593011 | 2 | 242,593,011 | ATG4B | A/G | 0.033 | 0.006 | 0/9/127 | 0/7/618 | 6.08 | 5.82E-04 |
| rs877859 | 3 | 107,714,075 | CD47 | G/A | 0.412 | 0.297 | 23/66/47 | 52/267/306 | 1.66 | 4.55E-04 |
| rs11921691 | 3 | 113,673,125 | ZDHHC23 | A/G | 0.375 | 0.506 | 22/58/56 | 168/297/160 | 0.62 | 6.26E-04 |
| rs6883840 | 5 | 40,286,410 | PTGER4 | A/G | 0.129 | 0.240 | 2/31/103 | 31/238/356 | 0.47 | 3.48E-05 |
| rs10057851 | 5 | 64,565,261 | ADAMTS6 | G/A | 0.375 | 0.502 | 13/76/47 | 153/321/151 | 0.60 | 5.24E-04 |
| rs78649652 | 5 | 96,124,373 | ERAP1 | A/G | 0.040 | 0.006 | 0/11/125 | 0/8/617 | 6.54 | 9.59E-05 |
| rs2122554 | 5 | 165,957,086 | ODZ2 | A/C | 0.081 | 0.031 | 2/18/116 | 1/37/587 | 2.73 | 4.93E-04 |
| rs11955074 | 5 | 178,294,060 | ZNF354B | A/G | 0.184 | 0.108 | 4/42/90 | 9/117/499 | 1.86 | 9.75E-04 |
| rs4712653 | 6 | 22,125,964 | LINC00340 | G/A | 0.559 | 0.452 | 40/72/24 | 122/321/182 | 1.64 | 5.73E-04 |
| rs6939340 | 6 | 22,140,004 | LINC00340 | G/A | 0.581 | 0.464 | 45/68/23 | 134/312/179 | 1.67 | 2.99E-04 |
| rs3095250 | 6 | 31,208,340 | HLA-C | G/A | 0.353 | 0.437 | 21/54/61 | 114/318/193 | 0.53 | 6.99E-04 |
| rs3130688 | 6 | 31,210,216 | HLA-C | G/A | 0.353 | 0.435 | 21/54/61 | 113/318/194 | 0.54 | 9.98E-04 |
| rs10274334 | 7 | 47,925,331 | PKD1L1 | G/C | 0.500 | 0.382 | 31/72/31 | 95/287/243 | 1.62 | 5.25E-04 |
| rs2247572 | 8 | 73,633,028 | KCNB2 | A/G | 0.206 | 0.12 | 5/46/85 | 6/138/481 | 1.90 | 3.89E-04 |
| rs3133745 | 8 | 96,534,806 | LOC100616530 | A/G | 0.206 | 0.125 | 5/46/85 | 11/134/480 | 1.82 | 9.45E-04 |
| rs34075341 | 9 | 91,616,843 | S1PR3 | A/G | 0.004 | 0.045 | 0/1/135 | 0/56/569 | 0.08 | 2.82E-04 |
| rs2418135 | 9 | 113,901,309 | OR2K2 | A/G | 0.382 | 0.520 | 19/66/51 | 172/306/147 | 0.55 | 2.20E-05 |
| rs56224008 | 9 | 131,107,634 | SLC27A4 | A/G | 0.066 | 0.023 | 0/18/118 | 1/27/597 | 2.98 | 7.09E-04 |
| rs61734605 | 11 | 34,916,657 | APIP | A/G | 0.422 | 0.307 | 25/65/46 | 60/264/301 | 1.75 | 1.10E-04 |
| rs938886 | 14 | 20,837,701 | TEP1 | G/C | 0.140 | 0.234 | 2/34/100 | 33/226/366 | 0.53 | 4.71E-04 |
| rs1713449 | 14 | 20,841,707 | TEP1 | A/G | 0.136 | 0.228 | 2/33/101 | 31/223/371 | 0.53 | 5.65E-04 |
| rs2069541 | 14 | 23,901,012 | MYH7 | G/A | 0.033 | 0.006 | 0/9/127 | 0/7/618 | 6.08 | 5.82E-04 |
| rs17101661 | 14 | 64,564,680 | SYNE2 | A/G | 0.048 | 0.011 | 0/13/123 | 0/14/611 | 4.43 | 2.66E-04 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs12918952 | 16 | 78,420,775 | WWOX | G/A | 0.294 | 0.423 | 15/50/71 | 110/309/206 | 0.58 | 2.99E-04 |
| rs79954845 | 17 | 36,483,889 | GPR179 | C/G | 0.030 | 0.003 | 2/4/130 | 0/4/621 | 9.44 | 2.42E-04 |
| rs16950981 | 18 | 6,992,683 | LAMA1 | A/T | 0.037 | 0.007 | 0/10/126 | 0/9/616 | 5.26 | 5.76E-04 |
| rs12961939 | 18 | 6,997,818 | LAMA1 | C/A | 0.191 | 0.290 | 4/44/88 | 59/245/321 | 0.58 | 7.12E-04 |
| bs19_9068458 | 19 | 9,068,458 | MUC16 | A/T | 0.026 | 0.002 | 0/7/129 | 0/3/622 | 10.98 | 3.99E-04 |

Table 3. Associations between genetic variants and PCa aggressiveness in the MI population with p-value < 1E-3.

| SNP | CHR | BP | Gene | A1/A2 | MAF Agg | MAF Non-Agg | Genotype Agg | Genotype Non-Agg | OR | P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| rs9701796 | 1 | 19,186,129 | TAS1R2 | C/G | 0.279 | 0.190 | 16/68/95 | 16/164/337 | 1.67 | 4.31E-04 |
| rs2272994 | 1 | 40,923,019 | ZNF643 | A/G | 0.288 | 0.193 | 10/83/86 | 20/159/338 | 1.72 | 2.06E-04 |
| rs28568406 | 1 | 158,687,163 | OR6K3 | A/G | 0.464 | 0.365 | 38/90/51 | 62/253/202 | 1.53 | 9.20E-04 |
| rs7530895 | 1 | 203,260,756 | LOC730227 | G/A | 0.008 | 0.044 | 0/3/176 | 1/43/473 | 0.19 | 6.72E-04 |
| rs669408 | 1 | 232,519,150 | SIPA1L2 | C/A | 0.489 | 0.378 | 44/85/48 | 79/233/205 | 1.54 | 4.46E-04 |
| rs2924461 | 5 | 8,012,069 | MTRR | G/A | 0.528 | 0.416 | 48/93/38 | 94/242/181 | 1.57 | 2.98E-04 |
| rs10499052 | 6 | 109,885,475 | AKD1 | A/G | 0.355 | 0.252 | 29/69/81 | 33/194/290 | 1.61 | 2.60E-04 |
| rs41289902 | 6 | 112,460,365 | LAMA4 | A/G | 0.028 | 0.004 | 0/10/169 | 0/4/513 | 7.40 | 4.09E-04 |
| rs117497357 | 7 | 20,768,077 | ABCB5 | A/T | 0.073 | 0.030 | 1/24/154 | 0/31/486 | 2.53 | 9.60E-04 |
| rs741301 | 7 | 36,917,995 | ELMO1 | G/A | 0.413 | 0.321 | 25/98/56 | 53/226/238 | 1.56 | 7.59E-04 |
| rs28382644 | 7 | 44,118,394 | POLM | G/C | 0.034 | 0.004 | 0/12/167 | 0/4/513 | 8.93 | 4.64E-05 |
| rs7787531 | 7 | 129,023,597 | AHCYL2 | G/A | 0.123 | 0.065 | 3/38/138 | 3/61/453 | 2.02 | 9.25E-04 |
| rs118014343 | 8 | 623,906 | ERICH1 | A/G | 0.056 | 0.019 | 0/20/159 | 0/20/497 | 3.00 | 7.95E-04 |
| rs7008113 | 8 | 111,438,655 | KCNV1 | A/G | 0.271 | 0.185 | 10/77/92 | 17/157/343 | 1.68 | 4.96E-04 |
| rs10090835 | 8 | 130,789,767 | GSDMC | A/G | 0.011 | 0.056 | 0/4/175 | 3/52/462 | 0.19 | 1.44E-04 |
| rs4919060 | 10 | 98,699,136 | LCOR | A/C | 0.221 | 0.148 | 8/63/108 | 10/133/374 | 1.74 | 6.61E-04 |
| rs61898615 | 11 | 103,019,260 | DYNC2H1 | A/G | 0.031 | 0.005 | 0/11/168 | 0/5/512 | 6.52 | 3.44E-04 |
| rs10841496 | 12 | 20,521,654 | PDE3A | A/C | 0.578 | 0.462 | 62/83/34 | 113/252/152 | 1.56 | 3.42E-04 |
| rs8176345 | 12 | 58,158,558 | CYP27B1 | A/G | 0.061 | 0.023 | 1/20/158 | 0/24/493 | 2.76 | 9.73E-04 |
| rs11116595 | 12 | 85,165,879 | SLC6A15 | A/G | 0.377 | 0.485 | 31/73/75 | 119/263/135 | 0.64 | 4.55E-04 |
| rs117476305 | 12 | 101,761,753 | UTP20 | G/A | 0.031 | 0.006 | 0/11/168 | 0/6/511 | 5.43 | 7.52E-04 |
| rs912969 | 13 | 103,867,104 | SLC10A2 | A/G | 0.028 | 0.076 | 0/10/169 | 3/73/441 | 0.35 | 6.72E-04 |
| rs4790404 | 17 | 2,886,642 | RAP1GAP2 | G/A | 0.391 | 0.502 | 27/86/66 | 135/248/133 | 0.63 | 2.23E-04 |
| rs1901187 | 17 | 38,646,147 | TNS4 | G/A | 0.346 | 0.458 | 19/86/74 | 99/276/142 | 0.61 | 1.85E-04 |
| rs7236632 | 18 | 55,434,202 | ATP8B1 | G/A | 0.224 | 0.139 | 9/62/108 | 7/130/380 | 1.83 | 1.87E-04 |
| rs34070230 | 19 | 4,844,790 | PLIN3 | C/G | 0.061 | 0.023 | 1/20/158 | 1/22/494 | 2.76 | 9.73E-04 |
| rs11881700 | 19 | 52,538,428 | ZNF432 | G/A | 0.187 | 0.115 | 5/57/117 | 7/105/405 | 1.77 | 8.20E-04 |

Table 4. Associations between genetic variants and PCa aggressiveness in the CAPS population with p-value < 1E-3.

| SNP | CHR | BP | Gene | A1/A2 | MAF Agg | MAF Non-Agg | Genotype Agg | Genotype Non-Agg | OR | P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| rs10797449 | 1 | 233,490,138 | KIAA1804 | A/G | 0.513 | 0.365 | 44/65/40 | 30/156/110 | 1.88 | 3.78E-05 |
| rs6757496 | 2 | 8,978,676 | KIDINS220 | A/G | 0.299 | 0.412 | 10/69/70 | 41/162/93 | 0.56 | 4.45E-04 |
| rs34276015 | 3 | 16,419,309 | RFTN1 | A/G | 0.007 | 0.049 | 0/2/147 | 0/29/267 | 0.13 | 6.82E-04 |
| rs13141997 | 4 | 99,478,756 | TSPAN5 | A/G | 0.440 | 0.319 | 26/79/44 | 27/135/134 | 1.73 | 4.37E-04 |
| rs1650697 | 5 | 79,950,781 | MSH3 | A/G | 0.188 | 0.291 | 2/52/95 | 23/126/147 | 0.57 | 8.49E-04 |
| rs17601580 | 6 | 132,061,420 | ENPP3 | A/G | 0.121 | 0.215 | 2/32/115 | 13/101/182 | 0.50 | 6.45E-04 |
| rs723874 | 8 | 20,107,601 | LZTS1 | C/G | 0.010 | 0.059 | 0/3/146 | 1/33/261 | 0.16 | 3.11E-04 |
| rs6475797 | 9 | 24,545,513 | ELAVL2 | G/A | 0.383 | 0.530 | 22/70/57 | 80/154/62 | 0.49 | 5.53E-06 |
| rs2071348 | 11 | 5,264,146 | HBBP1 | C/A | 0.242 | 0.372 | 8/56/85 | 36/148/112 | 0.50 | 5.53E-05 |
| rs10895391 | 11 | 103,158,278 | DYNC2H1 | A/G | 0.453 | 0.334 | 25/85/39 | 31/135/129 | 1.71 | 5.32E-04 |
| rs111874833 | 12 | 70,953,189 | PTPRB | A/G | 0.034 | 0.003 | 0/10/139 | 0/2/294 | 10.24 | 5.22E-04 |
| rs117710037 | 13 | 115,091,073 | CHAMP1 | A/C | 0.007 | 0.051 | 0/2/147 | 1/28/267 | 0.13 | 4.08E-04 |
| rs35572669 | 14 | 77,141,290 | VASH1 | C/A | 0.493 | 0.367 | 40/67/42 | 36/145/115 | 1.73 | 2.34E-04 |
| rs117555414 | 16 | 57,113,168 | NLRC5 | A/G | 0.030 | 0.002 | 1/7/141 | 0/1/295 | 18.40 | 3.44E-04 |
| bs16_75269325 | 16 | 75,269,325 | BCAR1 | A/C | 0.050 | 0.012 | 1/13/135 | 0/7/289 | 4.43 | 9.15E-04 |
| bs16_75301838 | 16 | 75,301,838 | BCAR1 | C/G | 0.050 | 0.012 | 1/13/135 | 0/7/289 | 4.43 | 9.15E-04 |
| rs1559806 | 18 | 72,108,787 | FAM69C | A/G | 0.500 | 0.370 | 35/79/35 | 33/152/110 | 1.76 | 3.52E-04 |
| rs266849 | 19 | 51,349,090 | KLK3 | G/A | 0.185 | 0.100 | 4/47/98 | 2/55/239 | 2.05 | 6.14E-04 |

We therefore conducted association analysis in a pooled sample of all three populations. A total of 39 genetic variants achieved a P-value < 1E-3 in the pooled analysis. Among them, 31 variants had effects in the same direction in all three populations. We further examined the 31 variants and found that 11 variants showed significant association ($P < 0.05$) in at least two populations. The association results of the 11 variants in each population and the pooled analysis were presented in Table 2.

Among those 11 variants, a rare but recurrent missense genetic variant bs2_233990575 in INPP5D region had consistent effect on PCa aggressiveness among all three populations at a liberal P-value of 0.05 (Table 4). It is located at 233,990,575 bp of chromosome 2, the bs2_233990575 rare allele 'A' appeared only in aggressive PCa subjects, with a minor allele frequency of 0.008, 0.008 and 0.010 in JHH, MI and CAPS population, respectively. On contrast, this rare allele was not observed among the 625, 517 and 296 non-aggressive PCa subjects in JHH, MI and CAPS population respectively (Table 4).

We therefore examined the other genetic variants located in the INPP5D gene region. We found that another rare missense variant rs115393439, which was 17 bp upstream of bs2_233990575, was significantly associated with PCa aggressiveness in the JHH and CAPS populations, with P of 0.012 and 0.050, respectively (Table 4). In addition, rs115393439 was associated with PCa aggressiveness with a marginal $P = 0.055$ in the MI populaiton (Table 4). Similar with bs2_233990575, the minor allele frequency of rs115393139 was higher in aggressive PCa

subjects than in non-aggressive PCa cases, resulting in an odds ratio (OR) of 5.83, 8.73 and 3.54 in JHH, MI and CAPS, respectively (Table 4).

Table 5. Associations between 11 selected variants and prostate cancer aggressiveness in stage 1.

| Population | SNP | CHR | BP | Gene | A1/A2 | MAF Agg | MAF Non-Agg | Genotype Agg | Genotype Non-Agg | OR | P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| JHH | | | | | | | | | | | |
| | bs1_156907115 | 1 | 156,907,115 | ARHGEF11 | G/A | 0.408 | 0.346 | 17/77/42 | 80/272/273 | 1.38 | 2.30E-02 |
| | bs2_233990575 | 2 | 233,990,575 | INPP5D | A/G | 0.007 | 0.000 | 0/2/134 | 0/0/625 | | 3.18E-02 |
| | rs115393439 | 2 | 233,990,592 | INPP5D | C/A | 0.018 | 0.003 | 0/5/131 | 0/4/621 | 5.83 | 1.19E-02 |
| | rs464494 | 5 | 76,003,258 | IQGAP2 | A/G | 0.375 | 0.457 | 22/58/56 | 120/331/174 | 0.70 | 1.42E-02 |
| | rs61740965 | 5 | 81,608,563 | ATP6AP1L | G/A | 0.022 | 0.004 | 0/6/130 | 0/5/620 | 5.62 | 6.40E-03 |
| | rs10274334 | 7 | 47,925,331 | PKD1L1 | G/C | 0.500 | 0.382 | 31/72/31 | 95/287/243 | 1.62 | 5.25E-04 |
| | rs7385804 | 7 | 100,235,970 | TFR2 | C/A | 0.456 | 0.378 | 27/70/39 | 91/290/244 | 1.42 | 1.35E-02 |
| | rs2418135 | 9 | 113,901,309 | OR2K2 | G/A | 0.618 | 0.480 | 19/66/51 | 172/306/147 | 1.82 | 2.20E-05 |
| | rs61753080 | 11 | 119,005,003 | HINFP | A/G | 0.007 | 0.004 | 0/2/133 | 0/5/616 | 1.85 | 3.64E-01 |
| | rs11116595 | 12 | 85,165,879 | SLC6A15 | A/G | 0.382 | 0.436 | 22/60/54 | 114/317/194 | 0.79 | 9.77E-02 |
| | rs17474506 | 17 | 38,990,780 | TMEM99 | G/C | 0.085 | 0.050 | 3/17/116 | 3/57/565 | 1.74 | 4.10E-02 |
| MI | | | | | | | | | | | |
| | bs1_156907115 | 1 | 156,907,115 | ARHGEF11 | G/A | 0.422 | 0.336 | 29/93/57 | 56/235/225 | 1.46 | 3.35E-03 |
| | bs2_233990575 | 2 | 233,990,575 | INPP5D | A/G | 0.008 | 0.000 | 0/3/176 | 0/0/517 | | 1.69E-02 |
| | rs115393439 | 2 | 233,990,592 | INPP5D | C/A | 0.008 | 0.001 | 0/3/176 | 0/1/516 | 8.73 | 5.47E-02 |
| | rs464494 | 5 | 76,003,258 | IQGAP2 | A/G | 0.358 | 0.460 | 20/88/71 | 113/250/154 | 0.66 | 1.16E-03 |
| | rs61740965 | 5 | 81,608,563 | ATP6AP1L | G/A | 0.017 | 0.004 | 0/6/173 | 0/4/513 | 4.39 | 2.23E-02 |
| | rs10274334 | 7 | 47,925,331 | PKD1L1 | G/C | 0.441 | 0.431 | 38/82/59 | 89/268/160 | 1.02 | 8.65E-01 |
| | rs7385804 | 7 | 100,235,970 | TFR2 | C/A | 0.455 | 0.374 | 42/79/58 | 85/217/215 | 1.34 | 1.40E-02 |
| | rs2418135 | 9 | 113,901,309 | OR2K2 | G/A | 0.500 | 0.480 | 45/89/45 | 116/263/137 | 1.09 | 4.71E-01 |
| | rs61753080 | 11 | 119,005,003 | HINFP | A/G | 0.020 | 0.005 | 0/7/170 | 0/5/510 | 4.14 | 1.58E-02 |
| | rs11116595 | 12 | 85,165,879 | SLC6A15 | A/G | 0.377 | 0.485 | 31/73/75 | 119/263/135 | 0.64 | 4.55E-04 |
| | rs17474506 | 17 | 38,990,780 | TMEM99 | G/C | 0.089 | 0.052 | 1/30/148 | 1/52/463 | 1.78 | 1.53E-02 |
| CAPS | | | | | | | | | | | |
| | bs1_156907115 | 1 | 156,907,115 | ARHGEF11 | G/A | 0.403 | 0.368 | 23/74/52 | 39/140/117 | 1.17 | 2.96E-01 |
| | bs2_233990575 | 2 | 233,990,575 | INPP5D | A/G | 0.010 | 0.000 | 0/3/146 | 0/0/296 | | 3.73E-02 |
| | rs115393439 | 2 | 233,990,592 | INPP5D | C/A | 0.023 | 0.007 | 0/7/142 | 0/4/292 | 3.54 | 4.96E-02 |
| | rs464494 | 5 | 76,003,258 | IQGAP2 | A/G | 0.366 | 0.411 | 19/71/59 | 47/149/100 | 0.84 | 2.50E-01 |
| | rs61740965 | 5 | 81,608,563 | ATP6AP1L | G/A | 0.003 | 0.002 | 0/1/148 | 0/1/295 | 1.99 | 1 |
| | rs10274334 | 7 | 47,925,331 | PKD1L1 | G/C | 0.534 | 0.444 | 45/69/35 | 60/143/93 | 1.46 | 8.54E-03 |
| | rs7385804 | 7 | 100,235,970 | TFR2 | C/A | 0.453 | 0.439 | 26/83/40 | 50/160/86 | 1.13 | 4.32E-01 |
| | rs2418135 | 9 | 113,901,309 | OR2K2 | G/A | 0.540 | 0.453 | 43/75/31 | 65/138/93 | 1.39 | 2.05E-02 |
| | rs61753080 | 11 | 119,005,003 | HINFP | A/G | 0.023 | 0.007 | 1/5/143 | 0/4/291 | 3.52 | 4.99E-02 |
| | rs11116595 | 12 | 85,165,879 | SLC6A15 | A/G | 0.389 | 0.476 | 25/66/58 | 63/155/77 | 0.72 | 2.81E-02 |
| | rs17474506 | 17 | 38,990,780 | TMEM99 | G/C | 0.074 | 0.052 | 1/20/128 | 1/29/266 | 1.44 | 2.30E-01 |
| Pooled | | | | | | | | | | | |
| | bs1_156907115 | 1 | 156,907,115 | ARHGEF11 | G/A | 0.412 | 0.347 | 69/244/151 | 175/646/615 | 1.33 | 3.68E-04 |
| | bs2_233990575 | 2 | 233,990,575 | INPP5D | A/G | 0.009 | 0.000 | 0/8/456 | 0/0/1437 | | 1.23E-05 |
| | rs115393439 | 2 | 233,990,592 | INPP5D | C/A | 0.016 | 0.003 | 0/15/449 | 0/9/1428 | 5.23 | 7.86E-05 |
| | rs464494 | 5 | 76,003,258 | IQGAP2 | A/G | 0.365 | 0.449 | 61/217/186 | 280/730/427 | 0.72 | 5.01E-05 |
| | rs61740965 | 5 | 81,608,563 | ATP6AP1L | G/A | 0.014 | 0.003 | 0/13/451 | 0/10/1427 | 4.07 | 9.41E-04 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs10274334 | 7 | 47,925,331 | PKD1L1 | G/C | 0.488 | 0.412 | 114/223/125 | 244/697/496 | 1.32 | 2.83E-04 |
| rs7385804 | 7 | 100,235,970 | TFR2 | C/A | 0.455 | 0.389 | 95/232/137 | 226/666/545 | 1.29 | 9.36E-04 |
| rs2418135 | 9 | 113,901,309 | OR2K2 | G/A | 0.547 | 0.474 | 139/230/95 | 328/706/402 | 1.37 | 4.81E-05 |
| rs61753080 | 11 | 119,005,003 | HINFP | A/G | 0.017 | 0.005 | 1/14/446 | 0/14/1416 | 3.59 | 7.98E-04 |
| rs11116595 | 12 | 85,165,879 | SLC6A15 | A/G | 0.383 | 0.462 | 78/199/187 | 296/735/406 | 0.72 | 2.92E-05 |
| rs17474506 | 17 | 38,990,780 | TMEM99 | G/C | 0.083 | 0.052 | 5/67/392 | 5/138/1293 | 1.67 | 7.34E-04 |

Gene-based association Analysis.

In addition to single variant analysis, we performed the gene-based association analysis in each population using SKAT. All polymorphic variants that passed quality control were included in the analysis. We found there were four genes, three genes and one gene significantly associated with PCa aggressiveness (p-value < 1E-4) in the JHH, MI and CAPS populations, respectively (Table 6). In the JHH population, the genes CREB3L1 (cAMP Responsive Element Binding Protein 3-like 1), KLF13 (Kruppel-like Factor 13), ROBO4 (Roundabout, Axon Guidance Receptor, Homolog 4), and ZCCHC6 (Zinc Finger, CCHC Domain Containing 6) presented significant association; in the Michigan population, the significant genes were TEK (Tyrosine Kinase, Endothelial), CDH2 (Cadherin 2), and BEST2 (Bestrophin 2); while in the CAPS population, the gene that showed significant association with PCa aggressiveness was actually a pseudogene LOC100128542. We then explored if there were genes contributing to PCa aggressiveness in at least two populations, setting a p-value threshold of 1E-3. We found 30 genes significantly associated with PCa aggressiveness in the JHH population (p-value < 1E-3), 22 genes in the MI population, and 70 genes in the CAPS population (Table 6). However, none of these genes that were significant at a P-value of 1E-03 were shared in more than 1 population.

Table 6. Gene-based analysis in JHH, MI and CAPS using SKAT.

| Population | SetID | P.value | N.Marker.All | N.Marker.Test |
|---|---|---|---|---|
| JHH | | | | |
| | CREB3L1 | 2.55E-06 | 9 | 9 |
| | KLF13 | 1.43E-05 | 1 | 1 |
| | ROBO4 | 2.63E-05 | 9 | 9 |
| | ZCCHC6 | 4.17E-05 | 7 | 7 |
| | RNF208 | 1.01E-04 | 2 | 2 |
| | LOC152742 | 1.24E-04 | 2 | 2 |
| | TRIM17 | 1.72E-04 | 3 | 3 |
| | L3MBTL2 | 1.78E-04 | 4 | 4 |

| | | | | |
|---|---|---|---|---|
| | SNX10 | 2.01E-04 | 2 | 2 |
| | CXorf68 | 2.01E-04 | 1 | 1 |
| | ZSCAN23 | 2.01E-04 | 1 | 1 |
| | F8 | 2.01E-04 | 2 | 2 |
| | FAM45A | 2.28E-04 | 1 | 1 |
| | KRTAP22-1 | 2.63E-04 | 2 | 2 |
| | RSG1 | 2.88E-04 | 3 | 3 |
| | TMEM177 | 3.11E-04 | 7 | 7 |
| | CDH6 | 3.32E-04 | 4 | 4 |
| | SPAG7 | 3.40E-04 | 2 | 2 |
| | RAB26 | 3.48E-04 | 3 | 3 |
| | IL16 | 3.70E-04 | 12 | 12 |
| | ZNF829 | 4.24E-04 | 3 | 3 |
| | EXOC3L2 | 4.32E-04 | 2 | 2 |
| | RIMS3 | 5.03E-04 | 3 | 3 |
| | MIR4697 | 5.60E-04 | 1 | 1 |
| | ARHGEF10 | 6.36E-04 | 12 | 12 |
| | C9orf135 | 6.41E-04 | 8 | 8 |
| | MLXIPL | 6.65E-04 | 6 | 6 |
| | PNMA2 | 6.73E-04 | 3 | 3 |
| | CCL16 | 9.79E-04 | 1 | 1 |
| | AHNAK | 9.98E-04 | 32 | 32 |
| MI | | | | |
| | TEK | 1.47E-05 | 9 | 9 |
| | CDH2 | 5.24E-05 | 14 | 14 |
| | BEST2 | 7.97E-05 | 4 | 4 |
| | LOC100130581 | 1.45E-04 | 1 | 1 |

| | | | | |
|---|---|---|---|---|
| | OR11L1 | 2.38E-04 | 8 | 8 |
| | S100PBP | 2.59E-04 | 4 | 4 |
| | LOC643339 | 4.25E-04 | 2 | 2 |
| | INMT | 5.07E-04 | 11 | 11 |
| | LOC148145 | 5.51E-04 | 1 | 1 |
| | NRIP3 | 5.80E-04 | 3 | 3 |
| | PPARGC1B | 6.33E-04 | 13 | 13 |
| | TIMM44 | 7.05E-04 | 8 | 8 |
| | LOC401164 | 7.18E-04 | 3 | 3 |
| | RFC1 | 7.47E-04 | 2 | 2 |
| | SLC16A5 | 7.47E-04 | 2 | 2 |
| | SRSF1 | 7.71E-04 | 1 | 1 |
| | MORN3 | 7.79E-04 | 4 | 4 |
| | CDH3 | 7.79E-04 | 10 | 10 |
| | DDHD1 | 8.71E-04 | 2 | 2 |
| | TMEM106C | 9.02E-04 | 5 | 5 |
| | KLK15 | 9.52E-04 | 4 | 4 |
| | LINC00284 | 9.80E-04 | 1 | 1 |
| CAPS | | | | |
| | LOC100128542 | 6.03E-05 | 2 | 2 |
| | HS3ST2 | 2.04E-04 | 2 | 2 |
| | PCDH8 | 2.47E-04 | 4 | 4 |
| | MRPL9 | 2.55E-04 | 7 | 7 |
| | PIGA | 2.85E-04 | 1 | 1 |
| | CCNI2 | 3.29E-04 | 1 | 1 |
| | MGC45800 | 3.40E-04 | 3 | 3 |
| | ZNF624 | 3.79E-04 | 2 | 2 |

| | | | |
|---|---|---|---|
| MAD2L1BP | 5.16E-04 | 1 | 1 |
| KIAA1462 | 5.41E-04 | 9 | 9 |
| BTN2A1 | 5.95E-04 | 4 | 4 |
| LOC645206 | 9.17E-04 | 2 | 2 |
| WDR72 | 9.19E-04 | 13 | 13 |

In addition to single variant and gene based association analysis, we also assessed the potential effects for all coding nonsynonymous variants by PolyPhen2. The possible impact of an amino acid substitution on the structure and function of a human protein was appraised quantitatively as benign, possibly damaging, or probably damaging. In table 7, we listed the prediction for the 7 missense variants of the 11 significant variants associated with PCa aggressiveness. According to PolyPhen2, 3 variants were predicted with benign effect, 2 with possibly damaging and 2 with probably damaging.

Table 7. Effect prediction for 11 significant variants associated with PCa aggressiveness by PolyPhen2.

| SNP | CHR | BP | Gene | A1/A2 | Annotation | Amino Acid Change | PolyPhen2 Prediction |
|---|---|---|---|---|---|---|---|
| bs1_156907115 | 1 | 156,907,115 | ARHGEF11 | G/A | missense | S(AGC) to G(GGC) | benign |
| bs2_233990575 | 2 | 233,990,575 | INPP5D | A/G | missense | R(CGC) to H (CAC) | possibly damaging |
| rs115393439 | 2 | 233,990,592 | INPP5D | C/A | missense | T(ACA) to P(CCA) | probably damaging |
| rs464494 | 5 | 76,003,258 | IQGAP2 | A/G | utr3 | | |
| rs61740965 | 5 | 81,608,563 | ATP6AP1L | G/A | missense | Y(TAC) to H(CAC) | probably damaging |
| rs10274334 | 7 | 47,925,331 | PKD1L1 | G/C | missense | R(CGC) to P(CCC) | benign |
| rs7385804 | 7 | 100,235,970 | TFR2 | C/A | Intron | | |
| rs2418135 | 9 | 113,901,309 | OR2K2 | G/A | Intergenic | | |
| rs61753080 | 11 | 119,005,003 | HINFP | A/G | missense | G(GGG) to E(GAG) | benign |
| rs11116595 | 12 | 85,165,879 | SLC6A15 | A/G | Intergenic | | |
| rs17474506 | 17 | 38,990,780 | TMEM99 | G/C | missense | I(ATC) to M(ATG) | possibly damaging |

## Discussion

To our knowledge, our study represents one of the first comprehensive studies to identify rare variants that are associated with aggressive PCa. Our data generated from the first 12-month funding period identified novel rare variants that are associated with aggressive PCa in Caucasians. We plan to conduct confirmation studies for the 11 significant variants in additional Caucasians and African American men.

We selected the Illumina Human Exome BeadChip (ExomeArray) as our genotyping platform to study rare variants. The ExomeArray chip represents the newest gene chip that delivers unparalleled coverage of putative functional exonic variants. The relatively cheaper cost makes it possible to study larger sample sizes. The Exome Beachip is comprised of >240,000 markers,

including >200,000 nonsynonymous SNPs, nonsense mutations, SNPs in splice sites and promoter regions, as well as thousands of GWAS tag markers. Nearly 90% of the SNPs on the exome arrays are rare, with a MAF<5%. In addition, the markers on the Illumina Human Exome BeadChips are selected from over 12,000 individual exome and whole-genome sequences, representing diverse populations, including those of European and African descent. Therefore, it is more efficient and economical to use exome arrays to identify rare variants associated with aggressive PCa, compared with whole genome sequencing.

As presented above, we found that two rare but recurrent variants that were 17 bp apart, in the INPP5D gene, were significantly associated with PCa aggressiveness. The INPP5D gene (Inositol Polyphosphate-5-Phosphatase 1) is a member of the INPP5 family. It encodes a Phosphatidylinositol (PtdIns) phosphatase that specifically hydrolyzes the 5-phosphate of phosphatidylinositol (3,4,5)-triphosphate (PtdIns (3,4,5) P3) to produce PtdIns(3,4)P2, and therefore negatively regulating the PI3K (phosphoinositide 3-kinase) pathways (Dunant et al. 2000). Acting as an inhibitor of the PI3K pathway, INPP5D is considered as a tumor suppressor in acute myeloid leukemia, Hodgkin's lymphoma, and acute lymphoblastic leukemia (Luo et al. 2004; Metzner et al. 2009; Tiacci et al. 2012). Besides, INPP5D has been found as the target of the cellular tumor antigen p53 in human breast cancer adenocarcinoma MCF7 cells and testicular germ cell tumor-derived human embryonal carcinoma cells (Kerley-Hamilton et al. 2005; Lion et al. 2013). Although the role INPP5D plays in prostate tumor cells has not been established, it is possible that INPP5D contributes to prostate cancer progression through the PI3K or p53 pathway.

Besides the single variant analysis, we also performed gene-based approach to identify genes that were associated with PCa aggressiveness. The gene-based approach (SKAT) we adopted is a novel statistical approach. SKAT is a supervised and flexible regression method to test for association between rare variants in a gene or genetic region and a continuous or dichotomous trait. Compared to other methods of estimating the joint effect of a subset of SNPs, SKAT is able to deal with variants that have different direction and magnitude of effects, and allows for covariate adjustment (Wu 2011). In addition, SKAT can also avoid arbitrary selection of threshold in burden test. Moreover, SKAT is computationally efficient, compared to a permutation test, making it feasible to analyze the large dataset in our study. Interestingly, several of the top targets identified by SKAT analysis (CREB3L1 and KLF13) encode transcription factors.

Besides all the above findings, we have also carefully calculated the study power based on our modified study design. We have >80% power to detect an OR of 1.7 (2.8) for variants with a MAF of 0.05 (0.01), at an alpha level of 1E-05 (2-sided). Therefore, we have sufficient power to identify novel rare mutations with relatively large effect based on our proposed sample size. We also considered several procedures to control for multiple test correction and SNP selection to be confirmed in additional independent samples. The Bonferroni corrected P-values are 2E-7 (0.05/200,000 variants) and 2E-6 (0.05/20,000 genes), for single variant analysis and gene-based

analysis, respectively. However, not all the tests for single variants are independent due to linkage disequilibrium (LD) structure among variants. In addition, previous studies also showed that the true associations do not necessarily reach the stringent Bonferroni corrected P-value cutoffs. Therefore, to balance study power and false positives, rare variants in Aim 1 that meet either of the following criteria with less stringent P-value cutoffs will be selected for replication: 1) variants reach a pooled analysis p-value of 1E-3 in single variant analysis; 2) variants with same effect direction in the JHH, Michigan and CAPS population; and 3) variants reached a p-value of 0.05 in at least two of the three populations. The adoption of the two-stage study design will further help to remove false positives.

In conclusion, we have identified several novel rare variants and genes that are associated with aggressive PCa in Caucasians. We expect that some of these significant variants could be confirmed in Caucasians and African American men in the next 12-month funding period. In that condition, the newly identified variants can provide more insight into the etiology of aggressive PCa and provide potential effective targets for therapy of aggressive PCa.

## KEY RESEARCH ACCOMPLISHMENTS

1) Completed IRB and other logistical issues
2) Performed single rare variant analysis, bioinformatics analysis, and gene-based analysis (SKAT) to identify rare variants that have strong effects on aggressive PCa risk in exome-array data among a total of 1,902 PCa cases, including 464 aggressive PCa cases and 1,438 indolent PCa cases.
3) Successfully identified 11 novel variants associated with PCa aggressiveness in Caucasians and further confirmation in additional Caucasians and African American men are to be conducted.

## CONCLUSION

1) We have made great progress in achieving the goals described in the approved Statement of Work.
2) We have identified 11 variants that are associated with aggressive PCa in Caucasians.
3) We plan to replicate the rare variants identified in Aim1 in additional 1,000 aggressive and 1,000 indolent PCa patients of European descent from JHH using Sequenom iPLEX MassARRAY platform.
4) We plan to evaluate the effect of rare mutations confirmed in Aim2 in an African American (AA) population with 500 aggressive and 500 indolent PCa patients from JHH.

5) We expect some of the variants identified in the first stage could be further confirmed, and therefore provide more insight into the etiology of aggressive PCa and provide potential effective targets for therapy of aggressive PCa.

## REPORTABLE OUTCOMES

1) Top variants in the genome that are significantly associated with aggressive PCa in EAs (Table 2 - Table 5)
2) Top genes in the genome that are significantly associated with aggressive PCa in EAs (Table 6)

## REFERENCES


Adzhubei I.A., Schmidt S., Peshkin L., et al. A method and server for predicting damaging missense mutations. Nat Methods. 7, 248-249 (2010).

Akbari MR, Trachtenberg J, Lee J, Tam S, Bristow R, Loblaw A, Narod SA, Nam RK. Association Between Germline HOXB13 G84E Mutation and Risk of Prostate Cancer. J Natl Cancer Inst. 2012 Aug 1;104(16):1260-2. Epub 2012 Jul 9.

Castro E. G.C.L., Olmos D., et al. Correlation of germ-line BRCA2 mutations with aggressive prostate cancer and outcome. ASCO Meet Abstr. 29, 1517 (2011).

Dunant NM, Wisniewski D, Strife A, Clarkson B, Resh MD (2000) The phosphatidylinositol polyphosphate 5-phosphatase SHIP1 associates with the dok1 phosphoprotein in bcr-Abl transformed cells. Cell Signal 12: 317-26.

Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, Wiley KE, Isaacs SD, Johng D, Wang Y, Bizon C, Yan G, Gielzak M, Partin AW, Shanmugam V, Izatt T, Sinari S, Craig DW, Zheng SL, Walsh PC, Montie JE, Xu J, Carpten JD, Isaacs WB, Cooney KA. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med. 2012 Jan 12;366(2):141-9.

Falk K, Rötzschke O. The final cut: how ERAP1 trims MHC ligands to size. Nat Immunol. 2002 Dec;3(12):1121-2.

Fredericks WJ, McGarvey T, Wang H, Lal P, Puthiyaveettil R, Tomaszewski J, Sepulveda J, Labelle E, Weiss JS, Nickerson ML, Kruth HS, Brandt W, Wessjohann LA, Malkowicz SB. The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression. DNA Cell Biol. 2011 Nov;30(11):851-64. Epub 2011 Jul 8.

Gallagher D.J., Gaudet M.M., Pal P., et al. Germline BRCA mutations denote a clinicopathologic subset of prostate cancer. Clin Cancer Res. 16, 2115-2121 (2010).

Hammer GE, Gonzalez F, Champsaur M, Cado D, Shastri N. The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules. Nat Immunol. 2006 Jan;7(1):103-12. Epub 2005 Nov 20.

Heemels MT, Ploegh H. Generation, translocation, and presentation of MHC class I-restricted peptides. Annu Rev Biochem. 1995;64:463-91. Review.

Henson BJ, Gollin SM. Overexpression of KLF13 and FGFR3 in oral cancer cells. Cytogenet Genome Res. 2010 Jun;128(4):192-8. doi: 10.1159/000308303. Epub 2010 Jun 2.

Hinks A, Martin P, Flynn E, Eyre S, Packham J; Childhood Arthritis Prospective Study-CAPS; BSPAR Study Group, Barton A, Worthington J, Thomson W. Subtype specific genetic associations for juvenile idiopathic arthritis: ERAP1 with the enthesitis related arthritis subtype and IL23R with juvenile psoriatic arthritis. Arthritis Res Ther. 2011 Jan 31;13(1):R12. doi: 10.1186/ar3235.

Kanazawa A, Kawamura Y, Sekine A, Iida A, Tsunoda T, Kashiwagi A, Tanaka Y, Babazono T, Matsuda M, Kawai K, Iiizumi T, Fujioka T, Imanishi M, Kaku K, Iwamoto Y, Kawamori R, Kikkawa R, Nakamura Y, Maeda S. Single nucleotide polymorphisms in the gene encoding Krüppel-like factor 7 are associated with type 2 diabetes. Diabetologia. 2005 Jul;48(7):1315-22. Epub 2005 Jun 4.

Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, Grönberg H, Wiklund F. A Population-based Assessment of Germline HOXB13 G84E Mutation and Prostate Cancer Risk. Eur Urol. 2012 Jul 20. [Epub ahead of print]

Kerley-Hamilton JS, Pike AM, Li N, DiRenzo J, Spinella MJ (2005) A p53-dominant transcriptional response to cisplatin in testicular germ cell tumor-derived human embryonal carcinoma. Oncogene 24: 6090-100. doi: 10.1038/sj.onc.1208755

Lion M, Bisio A, Tebaldi T, De Sanctis V, Menendez D, Resnick MA, Ciribilli Y, Inga A (2013) Interaction between p53 and estradiol pathways in transcriptional responses to chemotherapeutics. Cell Cycle 12: 1211-24. doi: 10.4161/cc.24309

Luo JM, Liu ZL, Hao HL, Wang FX, Dong ZR, Ohno R (2004) Mutation analysis of SHIP gene in acute leukemia. Zhongguo Shi Yan Xue Ye Xue Za Zhi 12: 420-6.

Mehta AM, Jordanova ES, Corver WE, van Wezel T, Uh HW, Kenter GG, Jan Fleuren G. Single nucleotide polymorphisms in antigen processing machinery component ERAP1 significantly associate with clinical outcome in cervical carcinoma. Genes Chromosomes Cancer. 2009 May;48(5):410-8.

Metzner A, Precht C, Fehse B, Fiedler W, Stocking C, Gunther A, Mayr GW, Jucker M (2009) Reduced proliferation of CD34(+) cells from patients with acute myeloid leukemia after gene transfer of INPP5D. Gene Ther 16: 570-3. doi: 10.1038/gt.2008.184

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904-9. Epub 2006 Jul 23.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics. 2007, 81.

Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, Greer F, Schomburg L, Fruci D, Niedermann G, van Endert PM. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat Immunol. 2005 Jul;6(7):689-97. Epub 2005 May 22.

Szczypiorska M, Sánchez A, Bartolomé N, Arteta D, Sanz J, Brito E, Fernández P, Collantes E, Martínez A, Tejedor D, Artieda M, Mulero J. ERAP1 polymorphisms and haplotypes are associated with ankylosing spondylitis susceptibility and functional severity in a Spanish population. Rheumatology (Oxford). 2011 Nov;50(11):1969-75. doi: 10.1093/rheumatology/ker229. Epub 2011 Aug 24.

Scohy S, Gabant P, Van Reeth T, Hertveldt V, Drèze PL, Van Vooren P, Rivière M, Szpirer J, Szpirer C. Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF transcription factor family. Genomics. 2000 Nov 15;70(1):93-101.

Tiacci E, Doring C, Brune V, van Noesel CJ, Klapper W, Mechtersheimer G, Falini B, Kuppers R, Hansmann ML (2012) Analyzing primary Hodgkin and Reed-Sternberg cells to capture the molecular and cellular pathogenesis of classical Hodgkin lymphoma. Blood 120: 4609-20. doi: 10.1182/blood-2012-05-428896

Thorne H., Willems A.J., Niedermayr E., et al. Decreased prostate cancer-specific survival of men with BRCA2 mutations from multiple breast cancer families. Cancer Prev Res (Phila). 4, 1002-1010 (2011).

Trembath RC Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2, Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, Barton A, Band G, Bellenguez C, Bergboer JG, Blackwell JM, Bramon E, Bumpstead SJ, Casas JP, Cork MJ, Corvin A, Deloukas P, Dilthey A, Duncanson A, Edkins S, Estivill X, Fitzgerald O, Freeman C, Giardina E, Gray E, Hofer A, Hüffmeier U, Hunt SE, Irvine AD, Jankowski J, Kirby B, Langford C, Lascorz J, Leman J, Leslie S, Mallbris L, Markus HS, Mathew CG, McLean WH, McManus R, Mössner R, Moutsianas L, Naluai AT, Nestle FO, Novelli G, Onoufriadis A,

Palmer CN, Perricone C, Pirinen M, Plomin R, Potter SC, Pujol RM, Rautanen A, Riveira-Munoz E, Ryan AW, Salmhofer W, Samuelsson L, Sawcer SJ, Schalkwijk J, Smith CH, Ståhle M, Su Z, Tazi-Ahnini R, Traupe H, Viswanathan AC, Warren RB, Weger W, Wolk K, Wood N, Worthington J, Young HS, Zeeuwen PL, Hayday A, Burden AD, Griffiths CE, Kere J, Reis A, McVean G, Evans DM, Brown MA, Barker JN, Peltonen L, Donnelly P, Trembath RC. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1.Nat Genet. 2010 Nov;42(11):985-90. doi: 10.1038/ng.694. Epub 2010 Oct 17.

Tryggvadottir L., Vidarsdottir L., Thorgeirsson T., et al. Prostate cancer progression and survival in BRCA2 mutation carriers. J Natl Cancer Inst. 99, 929-935 (2007).

Wu M.C., Lee S., Cai T., et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 89, 82-93 (2011).

Wan X, Yang C, Yang Q, Xue H, Tang NL, Yu W.Predictive rule inference for epistatic interaction detection in genome-wide association studies. Bioinformatics. 2010 Jan 1;26(1):30-7. Epub 2009 Oct 30.

Yan J, Parekh VV, Mendez-Fernandez Y, Olivares-Villagómez D, Dragovic S, Hill T, Roopenian DC, Joyce S, Van Kaer L. In vivo role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules. J Exp Med. 2006 Mar 20;203(3):647-59. Epub 2006 Feb 27.

Zhang,Y. and Liu,J.S. (2007) Bayesian inference of epistatic interactions in case-control studies. Nat. Genet., 39,1167–1173.

## TRAINING ACTIVITIES

My training was carried out following the training plan in the statement of work. The completed training activities include the several aspects: 1) reviewed literature on PCa genetics and genetic epidemiology: including but not limited to: clinical pathological characteristics of PCa, etiology of PCa, study designs in genetic epidemiology, on rare mutations and complex diseases; 2) received online education on ethics issues by Institutional Review Board (IRB) at WFUHS; 3) took two courses at WFUHS, including Introduction to Biostatistics (course code: CTPS730) and Epidemiology (CPTS720); 4) learned a set of software tools that are commonly used to manage and/or analyze genetic data, including PLINK, SKAT, EIGENSTRAT, STRUTURE, and SAS; 5) learned to use key bioinformatics tools, such as Polymorphism, and bioinformatics programming language, such as Perl and Python; 6) attended the 2013 American Society of Human Genetics meeting and earning continuing education credits; 7) attended a weekly journal club organized at the Center for Cancer Genomics, WFUSH.